Streamlined Romanization of Russian Cyrillic

Lyubomir Ivanov

Institute of Mathematics and Informatics
Bulgarian Academy of Sciences

Настоящая работа вводит Обтекаемую систему транслитерации русской кириллицы латинским алфавитом и сравнивает её с некоторыми существующими стандартами транслитерации. Статья использует опыт накопленный в процессе разработки и внедрения Обтекаемой системы транслитерации болгарской кириллицы, ставшей официальной для Болгарии, ООН, США и Великобритании.

The present work introduces the Streamlined System for the Romanization of Russian Cyrillic, and compares it with some of the existing transliteration standards. This paper draws from the experience gained with the development and implementation of the Streamlined System for the Romanization of Bulgarian Cyrillic, which is now official for Bulgaria, the UN, the USA and the UK.

Key words: transliteration, Russian Cyrillic, Streamlined System

1. Introduction

The streamlined approach to transliteration was initiated by the author with the development of the Streamlined System for the Romanization of Bulgarian. Originally, that system was introduced in Antarctica in 1995 (Ivanov 1995, Ivanov 2003), then in Bulgaria in 2000–2006 (Ivanov *et al.* 2010), and eventually codified by the Transliteration Act of 2009 (ДВ 2009). It was endorsed for official use by the United Nations in 2012 (UNGEGN 2013), and by the USA and the UK in 2013 (NGA 2014). The author proposes similar streamlined systems for the transliteration of other Slavic languages using Cyrillic script, notably the present one for Russian.

2. Latin transliteration (Romanization) of the Russian alphabet

There is a variety of different systems for the Romanization of the Russian alphabet (Beloozerov 2014). In particular, these include: the official Russian State Standard GOST R 7.0.34–2014 (ФАТРМ 2015); the ICAO system (ICAO

2015), used in Russian international passports, slightly modified to transliterate Russian hard sign **ъ** as **ie**; the BGN/PCGN 1947 system (USBGN 1994), for official US and UK use but also, with some modifications concerning Cyrillic **ё** and **ъ**, in Russian driving licenses; and the Library of Congress ALA-LC 1997 system (LOC 2012), used by American libraries, the British Library and some academic journals. That regulated official use goes along with an extensive practice of electronic communication in informally transliterated Russian (Zaliznyak and Mikaelyan 2006).

Far from becoming universal, the GOST R 7.0.34 standard has been largely disregarded with Russian and international institutions adopting other transliteration systems or turning to transcription instead, such as the IIHF 2010 system for transcribing the names of Russian ice hockey players (IIHF 2011). An attempt to introduce the GOST R 7.0.34 transliteration on Moscow street name plates puzzled Muscovites and foreign visitors alike (Streltsova et al. 2014) due to some of its features bordering on the bizarre, such as using Latin **x** and **shh** for Cyrillic **х** and **щ** respectively, and has apparently been abandoned.

In general, the present practice of Russian transliteration would seem fairly messy, inconsistent, and subject to not infrequent change.

3. The Streamlined System

The Streamlined System stands out among other approaches to the transliteration of Russian Cyrillic script in its sticking to transliteration proper (no transcription elements brought in), its being consistently English-oriented, taking advantage of the global *lingua franca* role of the English language (not only as a spoken language but also as a target language for the Romanization of a great number of languages including Chinese, Hindi, Japanese, Arabic etc.), and its seeking to strike an optimal balance between the following partly overlapping and partly conflicting priorities:

First, the primary purpose of this system is to ensure a plausible phonetic approximation of Russian words by English speaking users, including those having no knowledge of the Russian language and no available additional explanations;

Second – and of lesser priority, the system should allow for the retrieval of the original Cyrillic spellings as much as feasible;

Third, transliterated Russian words should fit an English language environment i.e. not be perceived as too 'un-English'; and

Fourth, transliterated word forms should be streamlined and simple. (Ivanov 2003, Ivanov *et al.* 2010)

Other possible considerations may relate to respecting established usage, avoiding undesirable associations in the case of overlapping forms of some words in different languages, or reflecting ideological motivation; see (Danchev

et al. 1989: 20–22) and (Selvelli 2015). These, however, are not prioritized highly by the streamlined approach.

Let us extend the standard Latin alphabet by adding the additional symbols ` (grave accent, used as a distinct symbol rather than diacritical mark), **"** (vertical double quotation mark) and **'** (vertical apostrophe). Then the Optimized Streamlined System for the Romanization of Russian alphabet (optimized with respect to the abovementioned priorities) is introduced by the following graphemic correspondences scheme:

> **а→a**, **б→b**, **в→v**, **г→g**, **д→d**, **е→e**, **ё→``e**, **ж→zh**,
> **з→z**, **и→i**, **й→y**, **к→k**, **л→l**, **м→m**, **н→n**, **о→o**,
> **п→p**, **р→r**, **с→s**, **т→t**, **у→u**, **ф→f**, **х→h**, **ц→ts**,
> **ч→ch**, **ш→sh**, **щ→shch**, **ъ→"** (or none), **ы→`i**,
> **ь** (before **е, ё, и, о, э**) → **y**, **ь** (otherwise) →**'** (or none),
> **э→`e**, **ю→yu**, **я→ya**

The Basic Streamlined System for Russian is defined by the above scheme with its additional symbols removed:

> **а→a**, **б→b**, **в→v**, **г→g**, **д→d**, **е→e**, **ё→e**, **ж→zh**,
> **з→z**, **и→i**, **й→y**, **к→k**, **л→l**, **м→m**, **н→n**, **о→o**,
> **п→p**, **р→r**, **с→s**, **т→t**, **у→u**, **ф→f**, **х→h**, **ц→ts**,
> **ч→ch**, **ш→sh**, **щ→shch**, **ъ →** (none), **ы→i**,
> **ь** (before **е, ё, и, о, э**) → **y**, **ь** (otherwise) → (none),
> **э→e**, **ю→yu**, **я→ya**

Typical for the streamlined approach is its non-use of diacritics, its use of Latin **y** for rendering only Cyrillic **й** rather than both **й** and **ы**, its non-use of Latin **j**, as well as its use of Latin **h** rather than **kh** for Cyrillic **х**.

In terms of compliance with the four priorities postulated above, the Optimized and the Basic version rate comparably with respect to the first one, while the former system complies better with the second one, and the latter has advantage with respect to the last two priorities.

4. Illustration

To illustrate the application of the streamlined approach, here follows some Russian text (excerpt from a novel by N.V. Gogol) transliterated both by the Optimized Streamlined System and by its basic version:

Russian Cyrillic

Эх, тройка! птица тройка, кто тебя выдумал? знать, у бойкого народа ты могла только родиться, в той земле, что не любит шутить, а ровнем-гладнем разметнулась на полсвета, да и ступай считать версты, пока

не зарябит тебе в очи. И не хитрый, кажись, дорожный снаряд, не железным схвачен винтом, а наскоро живьём с одним топором да долотом снарядил и собрал тебя ярославский расторопный мужик. Не в немецких ботфортах ямщик: борода да рукавицы, и сидит чёрт знает на чём; а привстал, да замахнулся, да затянул песню — кони вихрем, спицы в колесах смешались в один гладкий круг, только дрогнула дорога, да вскрикнул в испуге остановившийся пешеход — и вон она понеслась, понеслась, понеслась!

Н.В. Гоголь

Optimized Streamlined System

`Eh, troyka! ptitsa troyka, kto tebya v`idumal? znat', u boykogo naroda t`i mogla tol'ko rodit'sya, v toy zemle, chto ne lyubit shutit', a rovnem-gladnem razmetnulas' na polsveta, da i stupay schitay verst`i, poka ne zaryabit tebe v ochi. I ne hitr`iy, kazhis', dorozhn`iy snaryad, ne zhelezn`im shvachen vintom, a naskoro zhivy``em s odnim toporom da dolotom sobral tebya yaroslavskiy rastoropn`iy muzhik. Ne v nemetskih botfortah yamshchik: boroda da rukavits`i, i sidit ch``ert znaet na ch``em; a privstal, da zamahnulsya, da zatyanul pesnyu — koni vihrem, spits`i v kolesah smeshalis' v odin gladkiy krug, tol'ko drognula doroga, da vskriknul v ispuge ostanovivshiysya peshehod — i von ona poneslas', poneslas', poneslas'!

N.V. Gogol'

Basic Streamlined System

Eh, troyka! ptitsa troyka, kto tebya vidumal? znat, u boykogo naroda ti mogla tolko roditsya, v toy zemle, chto ne lyubit shutit, a rovnem-gladnem razmetnulas na polsveta, da i stupay schitay versti, poka ne zaryabit tebe v ochi. I ne hitriy, kazhis, dorozhniy snaryad, ne zheleznim shvachen vintom, a naskoro zhivyem s odnim toporom da dolotom sobral tebya yaroslavskiy rastoropniy muzhik. Ne v nemetskih botfortah yamshchik: boroda da rukavitsi, i sidit chert znaet na chem; a privstal, da zamahnulsya, da zatyanul pesnyu — koni vihrem, spitsi v kolesah smeshalis v odin gladkiy krug, tolko drognula doroga, da vskriknul v ispuge ostanovivshiysya peshehod — i von ona poneslas, poneslas, poneslas!

N.V. Gogol

5. Comparisons

By way of comparison, we present in the following table some sample words from the above text as transliterated also by the GOST R7.0.34, BGN/PCGN and ICAO systems.

Russian Cyrillic	Optimized Streamlined	Basic Streamlined	GOST R 7.0.34	BGN/PCGN	ICAO
эх	`eh	eh	e`x	ekh	ekh
тройка	troyka	troyka	trojka	troyka	troika
птица	ptitsa	ptitsa	ptica	ptitsa	ptitsa
выдумал	v`idumal	vidumal	vy`dumal	vydumal	vydumal
любит	lyubit	lyubit	lyubit	lyubit	liubit
хитрый	hitr`iy	hitriy	xitry`j	hitryy	hitryi
схвачен	shvachen	shvachen	skhvachen	skhvachen	skhvachen
ямщик	yamshchik	yamshchik	yamshhik	yamshchik	iamshchik
чёрт	ch``ert	chert	chyort	chërt	chert
вихрем	vihrem	vihrem	vixrem	vikhrem	vikhrem
только	tol'ko	tolko	tol`ko	tol'ko	tolko

Notice that the Basic Streamlined and ICAO transliterations use letters only, without any additional symbols.

6. Reversibility

The Optimized Streamlined System for Russian is not reversible, for it transliterates the Cyrillic letters and combinations of letters **ж**, **ц**, **ш**, **щ**, **тш**, **тщ**, **ю**, **я**, **ье**, **ьё**, **ьи**, **ьо** and **ьэ** in the same way as **зх**, **тс**, **сх**, **шч**, **цх**, **цхч**, **йу**, **йа**, **йе**, **йё**, **йи**, **йо** and **йэ** respectively. However, an auxiliary reversible variant of the system could be appropriate for certain special purposes (such as bibliographical documentation) where the exact retrieval of Russian words from their Romanized forms is deemed a priority. Namely, one may take the additional symbol · (interpunct) — or a conveniently available keyboard symbol such as | (vertical bar), drop the option of skipping **"** and **'**, and transliterate **зх**, **тс**, **сх**, **шч**, **цх**, **цхч** as **z·h**, **t·s**, **s·h**, **sh·ch**, **ts·h**, **ts·hch** respectively, and **й** before **у**, **а**, **е**, **ё**, **и**, **о** and **э** as **y·**. The plain optimized transliteration word form could be retrieved from its reversible variant by removing the additional symbol ·.

Although some of the letter combinations considered here would be infrequent or indeed unusual in Russian texts, they might occur in words of non-Russian origins or in misspelled Russian words. Furthermore, taking care of arbitrary Cyrillic letter combinations does not unnecessarily complicate the application of the reversible system as the transliteration of a given Russian text only depends on those members of the list **зх**, **тс**, **сх**, **шч**, **цх**, **цхч**, **йу**, **йа**, **йе**, **йё**, **йи**, **йо**, **йэ** that actually occur in the text.

We use the sample words from the preceding section to illustrate some differences between the reversible variant of the Streamlined System and the ALA-LC system, also reversible.

Russian Cyrillic	Reversible Streamlined	ALA-LC
эх	`eh	ėkh
тройка	troyka	troĭka
птица	ptitsa	pti͡tsa
выдумал	v`idumal	vydumal
любит	lyubit	li͡ubit
хитрый	hitr`iy	khitryĭ
схвачен	s·hvachen	skhvachen
ямщик	yamshchik	i͡amshchik
чёрт	ch``ert	chërt
вихрем	vihrem	vikhrem
только	tol'ko	tol'ko

7. Conclusion

The Romanization of Russian Cyrillic is currently carried out by means of several different transliteration systems (and their numerous versions), as well as by a multitude of informal ad-hoc transliteration methods commonly known as 'translit'. Yet none of the existing systems has proved potentially capable of gaining wider acceptance. We believe that the Streamlined System for Russian introduced in the present work has certain advantages making it promising in that respect.

R e f e r e n c e s

Beloozerov 2014: Б е л о о з е р о в, В. Латинский алфавит для русских слов. – *Проблемы современного образования*, 2014 № 2. 89–95. Available online at http://www.pmedu.ru/res/2014_2_11.pdf (accessed 25 February 2017)

Danchev et al. 1989: D a n c h e v, A., M. H o l m a n, E. D i m o v a, M. S a v o v a. *An English Dictionary of Bulgarian Names: Spelling and Pronunciation*. Sofia, Nauka i Izkustvo Publishers, 1989.

ДВ 2009: Закон за транслитерацията. *Държавен вестник* бр. 19, 13 март 2009, 6–8. Available at http://dv.parliament.bg/DVWeb/showMaterialDV.jsp?idMat=16943 (accessed 25 February 2017).

ФАТРМ 2015: *Национальный стандарт Российской Федерации ГОСТ Р* 7.0.34–2014: Правила упрощённой транслитерации русского письма латинским алфавитом. Федеральное агентство по техническому регулированию и метрологии. Москва, Стандартинформ, 2015. Available online at http://elib.pstu.ru/docview/docs/2503.pdf (accessed 25 February 2017).

ICAO 2015: Transliteration of Cyrillic Characters. – In: *Machine Readable Travel Documents: Transliterations recommended for use by States*. International Civil Aviation Organization, 2015. Doc 9303 (Seventh Edition) Part 3. 33–34. Available online at http://www.icao.int/publications/Documents/9303_p3_cons_en.pdf (accessed 25 February 2017).

IIHF 2011: *IIHF Transcription of Russian*. International Ice Hockey Federation, February 2011. Available online at http://www.iihf.com/fileadmin/user_upload/PDF/TranscriptionRUS.pdf (accessed 25 February 2017).

Ivanov 1995: I v a n o v, L. *Toponymic Guidelines for Antarctica*. Antarctic Place-names Commission of Bulgaria. Sofia, 1995. Available online at https://en.wikisource.org/wiki/Toponymic_Guidelines_for_Antarctica (accessed 25 February 2017).

Ivanov 2003: I v a n o v, L. On the Romanization of Bulgarian and English. – *Contrastive Linguistics*, XXVIII/2. 109–118. Errata: id., XXIX/1. 157.

Ivanov *et al*. 2010: I v a n o v, L., D. S k o r d e v, D. D o b r e v. The New National Standard for the Romanization of Bulgarian. – *Mathematica Balkanica*. New Series, Vol. 24, Fasc. 1-2. 121–130. Available online at http://www.fmi.uni-sofia.bg/fmi/logic/skordev/B1-Ivanov-Skordev-Dobrev.pdf (accessed 25 February 2017).

LOC 2012: *ALA-LC Romanization Tables: Russian*. The Library of Congress, 2012. Available online at https://www.loc.gov/catdir/cpso/romanization/russian.pdf (accessed 25 February 2017).

NGA 2014: *Romanization system for Bulgarian*: BGN/PCGN 2013 System. National Geospatial-Intelligence Agency, September 2014. Available online at http://geonames.nga.mil/gns/html/Romanization/Romanization_Bulgarian.pdf (accessed 25 February 2017).

Selvelli 2015: S e l v e l l i, G. Su alcuni aspetti ideologici dei sistemi di traslitterazione degli alfabeti cirillici nei Balcani. – *Studi Slavistici* XII. 159–180. Available online at http://www.fupress.net/index.php/ss/article/download/17975/16784 (accessed 25 February 2017).

Streltsova et al. 2014: С т р е л ь ц о в а, О., М. Г л е б о в а, А. Д е р з а е в а. *Адресные таблички в Москве путают иностранцев*. ТВ Центр, октябрь 2014. Available online at http://www.tvc.ru/news/show/id/53538/ (accessed 25 February 2017).

UNGEGN 2013: *Report on the Current Status of United Nations Romanization Systems for Geographical Names: Bulgarian*. Compiled by the UNGEGN Working Group on Romanization Systems. Version 4.0, February 2013. Available online at http://www.eki.ee/wgrs/rom1_bg.htm (accessed 25 February 2017).

USBGN 1994: Romanization System for Russian: BGN/PCGN 1947 System. – In: *Romanization Systems and Roman-Script Spelling Conventions*. Defense Mapping Agency, 1994. 93–94. Available online at http://libraries.ucsd.edu/bib/fed/USBGN_romanization.pdf (accessed 25 February 2017).

Zaliznyak and Mikaelyan 2006: З а л и з н я к, А., И. М и к а э л я н. Переписка по электронной почте как лингвистический объект. – In: *Труды международной конференции Диалог 2006 по компьютерной лингвистике и ее приложениям*. 157–162. Available online at http://www.dialog-21.ru/digests/dialog2006/materials/pdf/Zaliznyak.pdf (accessed 25 February 2017).

Обтекаема система за транслитерация на руската кирилица

Любомир Иванов

Институт по математика и информатика
Българска академия на науките

Настоящата работа въвежда Обтекаемата система за транслитерация на руската кирилица с латиница и я сравнява с някои съществуващи стандарти за транслитерация. Статията използва опита, натрупан в процеса на разработката и внедряването на Обтекаемата система за транслитерация на българската кирилица, станала официална за България, ООН, САЩ и Великобритания.

e-mail: lyubomail@yahoo.com
Institute of Mathematics and Informatics
Acad. Georgi Bonchev Street, Block 8
1113 Sofia